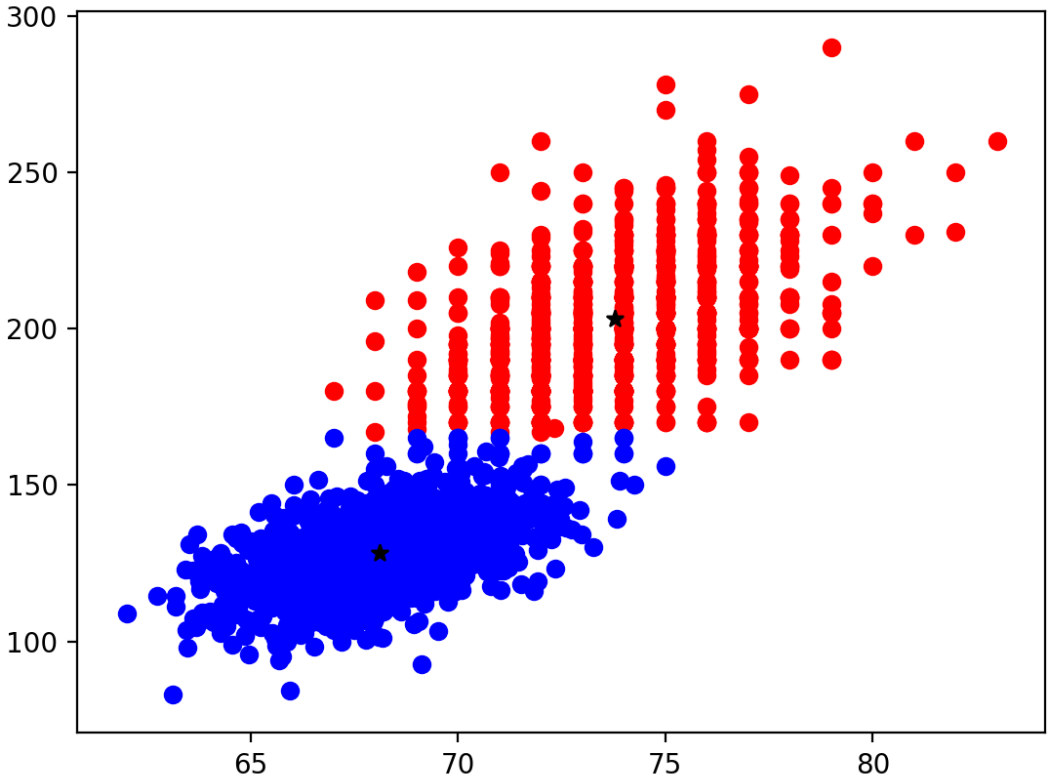
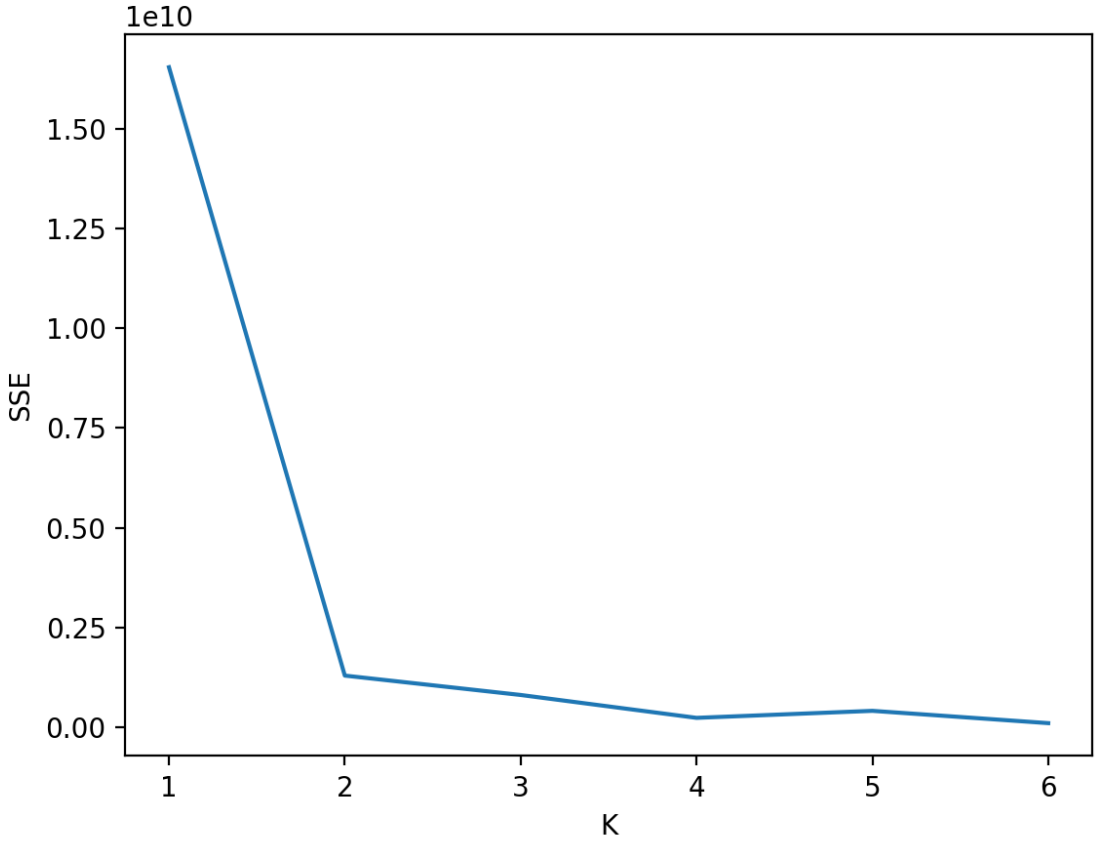
1) Unsupervised Learning with Clustering



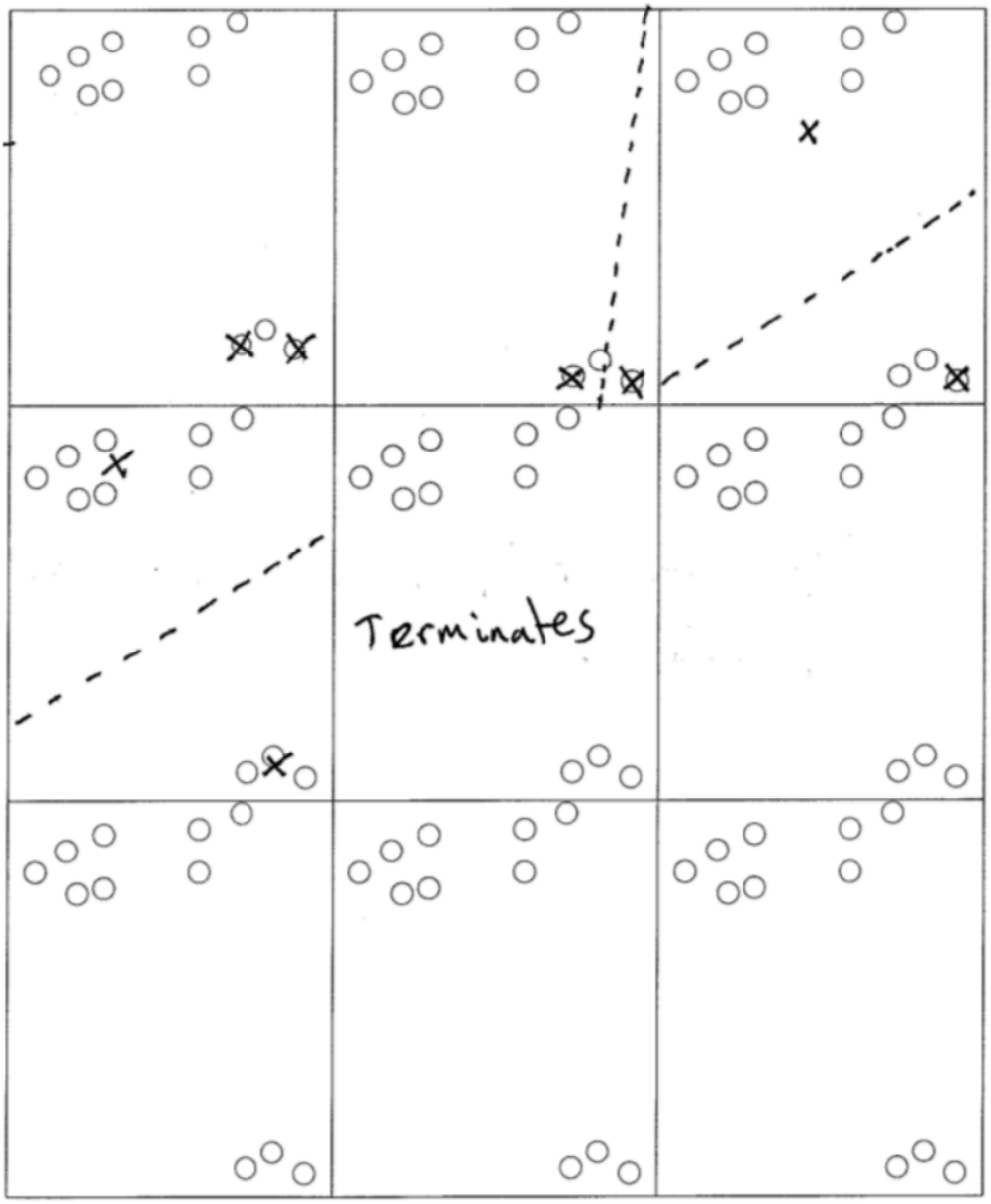


```
...Purities...
[0.9724249797242498, 0.999000999000999]
```

2) Sample Exam Questions

Question: 1. K-means and Gaussian Mixture Models

(a) Run k-means manually for the following dataset, where k = 2. Circles are data points and squares are the initial cluster centers. Draw the cluster centers and the decision boundaries that define each cluster. Use as many pictures as you need until convergence.



Question: 2. K-means Clustering

(a) How many 3-starting configurations are there? (Remember, a 3-starting configuration is just a size 3 subset of the 6 datapoints.

Answer: $\binom{6}{3} = \frac{6!}{3!*3!} = \frac{6*5*4}{3*2} = 20$

b) Fill in the table

3-partition	Is it stable?	An example 3-starting configuration that can arrive at the 3-partition after 0 or more iterations of k-means (or write "none" if no such 3-starting configuration)	The number of unique starting configurations that can arrive at the 3-partition.
$\{a,b,e\},\{c,d\},\{f\}$	N	none	0
$\{a,b\},\{d,e\},\{c,f\}$	Y	{b,c,e}	4
$\{a,d\},\{b,e\},\{c,f\}$	Y	{a,b,c}	8
$\{a\},\{d\},\{b,c,e,f\}$	Y	{a,b,d}	2
$\{a,b\},\{d\},\{c,e,f\}$	Y	none	0
$\{a,b,d\},\{c\},\{e,f\}$	Y	{a,c,f}	1

Question: 3. Decision Trees

(a) [2 points] What is H(Emotion|Wig = Y) (where H is entropy)?

Answer: $-(\frac{1}{2} * log(\frac{1}{2}) + \frac{1}{2} * log(\frac{1}{2})) = -2(-0.5) = 1$

(b) [2 points] What is H(Emotion|Ears = 3)?

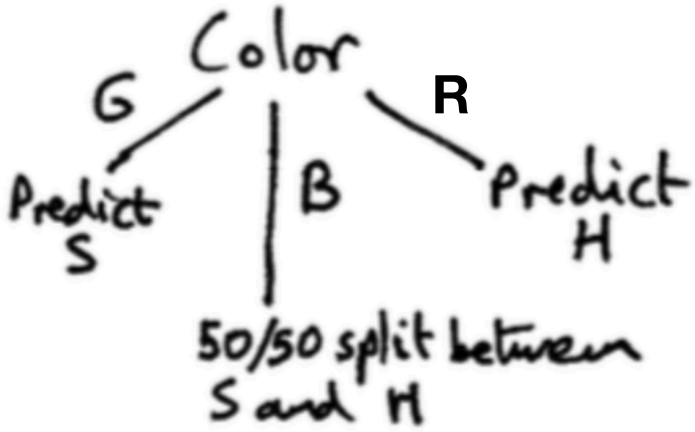
Answer: $-(1) * log(1) = 0$

(c) [3 points] Which attribute would the decision-tree building algorithm choose to use for the root of the tree (assume no pruning)?

Answer: Color would be the best attribute to choose for the root since it eliminates the need for us to ever use either of the other two attributes. From the given graph it's easy to see that choosing green will always result in sadness and choosing

red will always result in happiness. Blue will then be an even split between sad and happy.

(d) [3 points] Draw the full decision tree that would be learned from this data (assume no pruning).



Answer:

(e) [3 points] Assuming that the output attribute can take two values (i.e. has arity 2) what is the maximum training set error (expressed as a percentage) that any dataset could possibly have?

Answer: Since we are dealing with an output attribute that takes two values, our training set error will be at most 50%. In a case like this we are essentially selecting the output randomly.

(f) [3 points] Construct an example dataset that achieves this maximum percentage training set error (it must have two or fewer inputs and five or fewer records).

Answer: x: {1, 1, 0, 0} y: {0, 1, 0, 1}